parte01_analise_projecao_populacao

December 4, 2025

```python
[1]: import polars as pl
     import matplotlib.pyplot as plt
```

```python
[2]: df = pl.read_excel('dados/projecoes_2024_tab2_grupo_quinquenal.xlsx')
```

```python
[3]: df.head()
```

[3]: shape: (5, 76)

GRUPO ETÁRIO	SEXO	CÓD.	SIGLA	…	2067	2068	2069	2070
str	str	i64	str		i64	i64	i64	i64
00-04	Ambos	0	BR	…	7744772	7673320	7606268	7543093
05-09	Ambos	0	BR	…	8168636	8071186	7979360	7893433
10-14	Ambos	0	BR	…	8703349	8591634	8480688	8371817
15-19	Ambos	0	BR	…	9236029	9132638	9026379	8917283
20-24	Ambos	0	BR	…	9697332	9606213	9512994	9417186

```python
[4]: df = df.drop(['CÓD.', 'SIGLA'])
```

```python
[5]: df.select(pl.col('SEXO').unique())
```

[5]: shape: (3, 1)

SEXO

```
---
str

Mulheres
Ambos
Homens
```

[6]: `df.select(pl.col('LOCAL').unique())`

[6]: shape: (33, 1)

```
LOCAL
---
str

Alagoas
Rio Grande do Sul
Distrito Federal
Minas Gerais
Centro-Oeste
…
Rondônia
Piauí
Maranhão
Mato Grosso do Sul
Sul
```

[7]:
```python
df = df.filter([
    (pl.col('SEXO') == 'Ambos'),
    (pl.col('LOCAL') == 'Brasil')
])
```

[8]:
```python
print(df.select(pl.col('SEXO').unique()))
print(df.select(pl.col('LOCAL').unique()))
```

shape: (1, 1)

```
SEXO
---
str

Ambos
```

shape: (1, 1)

```
LOCAL
```

```
---
str

Brasil
```

```
[9]: print(df.columns)
```

```
['GRUPO ETÁRIO', 'SEXO', 'LOCAL', '2000', '2001', '2002', '2003', '2004',
 '2005', '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014',
 '2015', '2016', '2017', '2018', '2019', '2020', '2021', '2022', '2023', '2024',
 '2025', '2026', '2027', '2028', '2029', '2030', '2031', '2032', '2033', '2034',
 '2035', '2036', '2037', '2038', '2039', '2040', '2041', '2042', '2043', '2044',
 '2045', '2046', '2047', '2048', '2049', '2050', '2051', '2052', '2053', '2054',
 '2055', '2056', '2057', '2058', '2059', '2060', '2061', '2062', '2063', '2064',
 '2065', '2066', '2067', '2068', '2069', '2070']
```

```
[10]: df = df.drop(['SEXO', 'LOCAL'])
```

```
[11]: df = df.rename({'GRUPO ETÁRIO': 'grupo_etario'})
```

```
[12]: print(df.columns)
```

```
['grupo_etario', '2000', '2001', '2002', '2003', '2004', '2005', '2006', '2007',
 '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017',
 '2018', '2019', '2020', '2021', '2022', '2023', '2024', '2025', '2026', '2027',
 '2028', '2029', '2030', '2031', '2032', '2033', '2034', '2035', '2036', '2037',
 '2038', '2039', '2040', '2041', '2042', '2043', '2044', '2045', '2046', '2047',
 '2048', '2049', '2050', '2051', '2052', '2053', '2054', '2055', '2056', '2057',
 '2058', '2059', '2060', '2061', '2062', '2063', '2064', '2065', '2066', '2067',
 '2068', '2069', '2070']
```

```
[13]: df_tidy = df.unpivot(
          index='grupo_etario',
          on=[str(x) for x in range(2000, 2071)],
          variable_name='ano',
          value_name='populacao'
      )
```

```
[14]: df_tidy.head()
```

[14]: shape: (5, 3)

grupo_etario	ano	populacao
str	str	i64
00-04	2000	17281547
05-09	2000	17378990

```
    10-14        2000   17599378
    15-19        2000   17831344
    20-24        2000   16402228
```

[15]:
```python
df_grouped = df_tidy.group_by('ano').agg(
    pl.col('populacao').sum().alias('populacao_br')
)
```

[16]:
```python
df_grouped = df_grouped.with_columns(
    pl.col('ano').cast(int)
)
```

[17]:
```python
df_grouped.head()
```

[17]: shape: (5, 2)

```
    ano    populacao_br
    ---    ---
    i64    i64

    2000   174695935
    2005   185518369
    2007   189458827
    2013   199226702
    2036   219681891
```

[18]:
```python
df_pd = df_grouped.to_pandas()
```

[19]:
```python
df_pd.head()
```

[19]:
```
      ano   populacao_br
0    2000      174695935
1    2005      185518369
2    2007      189458827
3    2013      199226702
4    2036      219681891
```

[20]:
```python
df_pd = df_pd.sort_values('ano').reset_index(drop=True)
```

[21]:
```python
df_pd.head()
```

[21]:
```
      ano   populacao_br
0    2000      174695935
1    2001      177003743
2    2002      179228254
3    2003      181377654
```

```
4   2004      183469593
```

[22]: `df_pd.dtypes`

[22]:
```
ano               int64
populacao_br      int64
dtype: object
```
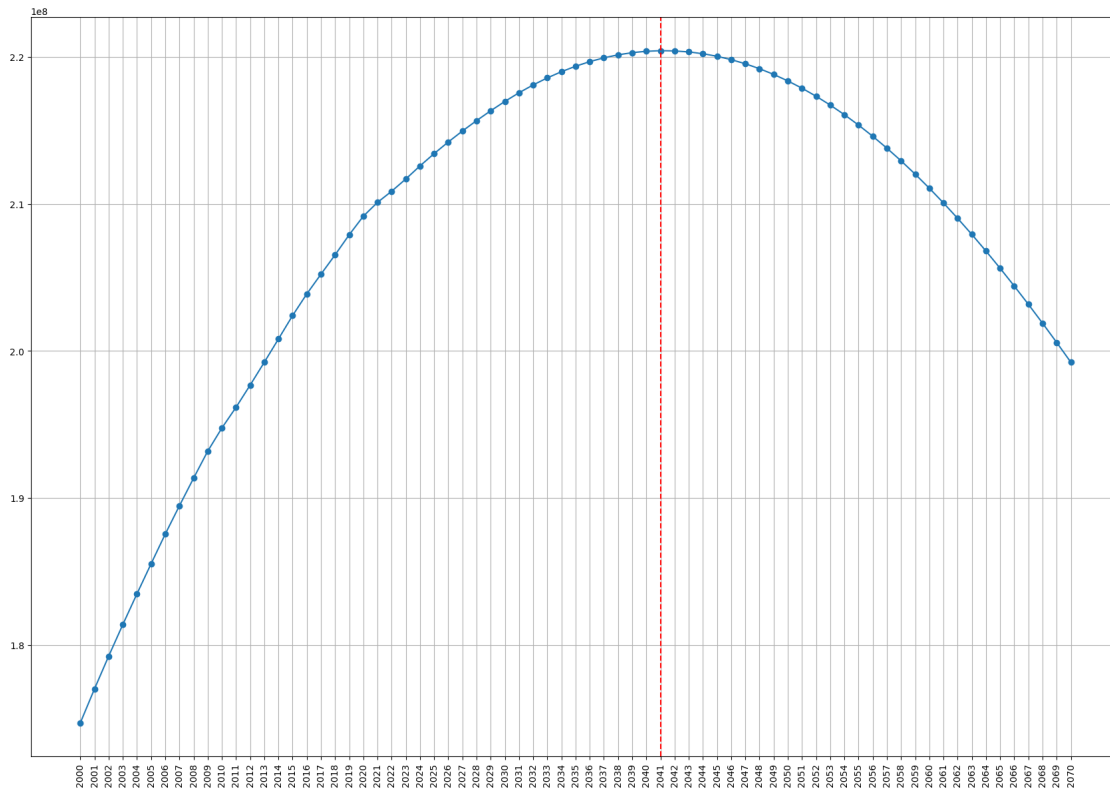
[23]:
```
num_habintes_pico = df_pd.query("populacao_br == populacao_br.max()").
 ↪reset_index()['ano'].sum()
```

[24]:
```
fig, ax = plt.subplots(figsize=(17, 12))

x = df_pd['ano']
y = df_pd['populacao_br']

ax.plot(x, y, marker='o')
ax.tick_params('x', rotation=90)
ax.grid()
ax.set_xticks(x.unique())
ax.axvline(num_habintes_pico, linestyle='--', color='red')
#plt.title('Projeção da população brasileira pelo IBGE')

plt.tight_layout()
fig.savefig('projecao_populacao_br.png', dpi=300, bbox_inches='tight')
plt.show()
```

```
[25]: df_pd.query("populacao_br == populacao_br.min()")
```

```
[25]:       ano  populacao_br
      0    2000     174695935
```

```
[26]: df_pd.query("populacao_br == populacao_br.max()")
```

```
[26]:       ano  populacao_br
      41   2041     220425299
```

```
[27]: df_pd.query("ano == 2070")
```

```
[27]:       ano  populacao_br
      70   2070     199228708
```

```
[28]: df_pd.query("ano == 2013")
```

```
[28]:       ano  populacao_br
      13   2013     199226702
```

```
[29]: ano_2070 = df_pd.query("ano == 2070")['populacao_br'].sum()
      ano_2013 = df_pd.query("ano == 2013")['populacao_br'].sum()
      ano_2070 - ano_2013
```

```
[29]: np.int64(2006)
```